SDAIA
الهيئة السعودية للبيانات والذكاء الاصطناعي
Saudi Data & AI Authority

مبـــادئ

الذكاء الاصطناعي التوليدي للعموم

2025م

محتويـــات

المقدمة	04
1. التعريفات	05
2. النطاق	05
3. فوائد الذكاء الاصطناعي التوليدي	06
4. مبادئ الذكاء الاصطناعي التوليدي	08
5. مخاطر الذكاء الاصطناعي التوليدي وطرق تخفيفها	13

المقدمة

شهد مجال الذكاء الاصطناعي التوليدي اهتماماً كبيراً في وسائل الإعلام في الأشهر الأخيرة وعلى الرغم من أنه جديد فقد بدأ كل من الأفراد والشركات بدمجه في حياتهم وأنشطتهم اليومية، لما تقدمه هذه التقنية من مزايا ومنافع، ومع ذلك من المهم إدراك المخاطر والتحديات المحتملة من إساءة الاستخدام.

تقوم الهيئة السعودية للبيانات والذكاء الاصطناعي (سدايا) بتقييم شامل لهذه التقنية بهدف تشجيع الاستثمار وتعزيز الاستخدام المسؤول لها. يمكن لأدوات الذكاء الاصطناعي التوليدي تحسين الكفاءة التشغيلية وإنشاء منتجات جديدة ليستفيد منها كافة فئات المجتمع بما يساعد على تحقيق الأهداف الوطنية في المملكة العربية السعودية.

1. التعريفات

لأغراض هذه الوثيقة يقصد بالمصطلحات والعبارات الآتية أينما وردت في هذه الوثيقة المعاني الموضحة أمام كل منها، ما لم يقتضِ السياق غير ذلك:

الذكاء الاصطناعي التوليدي: النموذج توليدي، نموذج تعلم الآلة يمكنه إنشاء أمثلة جديدة مشابهة لمجموعة بيانات التدريب، كما يعد هذا النموذج جزءاً فرعياً من الذكاء الاصطناعي يمكنه إنشاء محتوى جديد (بما في ذلك: النصوص والصور والصوت والرموز ومقاطع الفيديو وغيرها) ويعمل عن طريق تفسير الأوامر التي يقدمها المستخدمون، ويمكن أن يؤدي الذكاء الاصطناعي التوليدي أداء المهام التي تتطلب قدرات معرفية بشرية، بما في ذلك الاستجابة لصياغة الأوامر اللفظية أو المكتوبة، والتعلم، وحل المشكلات.

المستخدمون: أي شخص ذي صفة طبيعية أو اعتبارية يستهلك أو يستخدم السلع أو الخدمات التي تنتجها أنظمة الذكاء الاصطناعي التوليدي، ويشمل ذلك الشركات والجهات غير الحكومية والأفراد.

المطورون: أي شخص ذي صفة طبيعية أو اعتبارية يقوم بتطوير أنظمة الذكاء الاصطناعي التوليدي لتحقيق أهداف معينة، ويشمل ذلك مطوري الذكاء الاصطناعي وعلماء البيانات والباحثين.

2. النطاق

أعدت هذه الوثيقة **لتكون دليلاً للعموم بما يشمل مطوري ومستخدمي الذكاء الاصطناعي التوليدي في المملكة**

تعد هذه الوثيقة مرجعاً شاملاً يهدف إلى تشجيع تطوير التقنية بما يعزز الاستخدام المسؤول والفعال للذكاء الاصطناعي التوليدي في المملكة العربية السعودية، تنطبق هذه الوثيقة على العموم بما في ذلك من يقوم بتصميم أو تطوير أو نشر أو تنفيذ أو استخدام أو التأثر بأنظمة الذكاء الاصطناعي التوليدي. وعليه؛ قامت الهيئة السعودية للبيانات والذكاء الاصطناعي (سدايا) بصياغة هذه الوثيقة كمبادئ لتعزيز الفهم والثقة بتقنيات الذكاء الاصطناعي التوليدي.

3. فوائد الذكاء الاصطناعي التوليدي

من المتوقع أن يؤدي الذكاء الاصطناعي التوليدي إلى إحداث تطور لمختلف القطاعات داخل المملكة العربية السعودية بالطرق التالية:

3.1 زيادة الكفاءة والإنتاجية

يمكن للذكاء الاصطناعي التوليدي أتمتة مجموعة واسعة من المهام الروتينية، مثل: تلخيص الوثائق المعقدة، أو نسخ التسجيل الصوتي إلى نصوص، أو توليد مقاطع الفيديو والصور، مما يسمح بالتركيز على المهام الاستراتيجية والأكثر تعقيداً.

3.2 تحسين التواصل والتعاون

يساعد تبني الذكاء الاصطناعي التوليدي على تعزيز التواصل والتعاون في مجالات، مثل: توليد المحتوى، وإنشاء الصور الرقمية، وتحسين عرض النطاق الترددي للاجتماعات المرئية، وصياغة بيئات عمل افتراضية متطورة، وهذه العناصر يمكن أن تساهم في تعزيز الاتصال والتعاون.

3.3 اتخاذ القرارات المدروسة

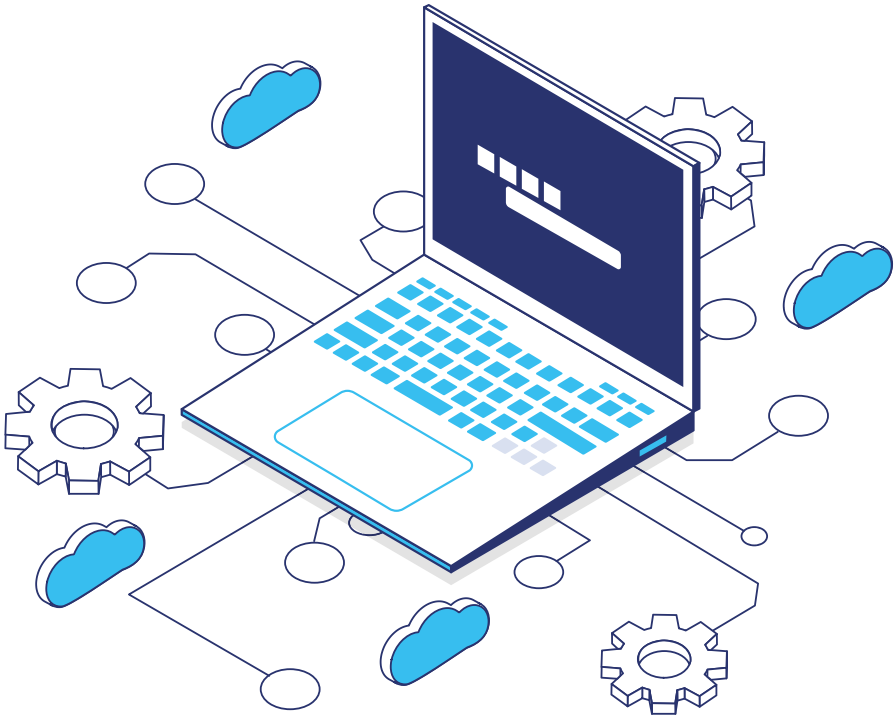
يمكن أن تساعد خدمات الذكاء الاصطناعي التوليدي على عملية اتخاذ القرار من خلال تحليل وتلخيص البيانات المتعددة التنسيقات (على سبيل المثال: النص والصوت والفيديو) والمساعدة على محاكاة السيناريوهات التي يمكن أن تؤثر على القرارات الاستراتيجية، وعند تطبيقها بشكل مناسب تعزز هذه التقنية النهج القائم على البيانات في صنع القرار في مجالات متنوعة، مثل: إدارة الأزمات، والرعاية الصحية، والتعليم.

3.4 تحسين إمكانية الوصول والشمولية

يقديقدم الذكاء الاصطناعي التوليدي حلولاً لتعزيز إمكانية الوصول إلى المحتوى، ويمكن أن ينتج مواد مخصصة للأفراد ذوي الإعاقة، مما يساهم في الوصول إلى المنصات والتواصل على مستوى العالم من خلال إنشاء محتوى يتماشى مع اللهجات والثقافات المحلية إذ يعزز التواصل بين الخلفيات المتنوعة من الأشخاص.

3.5 رفع جودة الخدمات العامة

يتمتع الذكاء الاصطناعي التوليدي بالقدرة على تعزيز رضا العملاء عن الخدمات العامة من خلال تقديم الخدمات حسب الحاجة وتوسيع توفرها، ويمكن استخدام التقنية لتحسين تجارب المواطنين في مختلف القطاعات بما في ذلك الرعاية الصحية والتعليم والرعاية الاجتماعية والمساعدة القانونية.



4. مبادئ الذكاء الاصطناعي التوليدي

عند التعامل مع أدوات الذكاء الاصطناعي التوليدي يجب على المطورين والمستخدمين الالتزام بأخلاقيات الذكاء الاصطناعي ومبادئ الذكاء الاصطناعي التوليدي خلال جميع مراحل دورة حياة هذه الأدوات لتسخير فوائده والتخفيف من المخاطر.

تنطبق مبادئ أخلاقيات الذكاء الاصطناعي التي طورتها سدايا على كافة فئات المجتمع في المملكة العربية السعودية، وتشمل –على سبيل المثال لا الحصر- الجهات العامة والخاصة وغير الربحية والباحثين والموظفين في القطاع العام والخاص والمستهلكين. كما صممت المبادئ لتنطبق على جميع أدوات الذكاء الاصطناعي المستخدمة وليس فقط الذكاء الاصطناعي التوليدي وهي مدرجة أدناه:

4.1 النزاهة والإنصاف

من الضروري اتخاذ الإجراءات اللازمة لتعريف الفئات المتأثرة بالنظام والتأكد من عدم وجود التحيز أو التمييز أو التنميط الذي يتعرض له الأفراد أو الجماعات أو الفئات بسبب البيانات أو الخوارزميات التي يمكن أن تؤدي إلى تمييز سلبي لفئة محددة، وتتمتع أدوات الذكاء الاصطناعي التوليدي بالقدرة على إنشاء محتوى قد يكون متحيزاً أو يفتقر إلى التمثيل، ولضمان نظم متسقة قائمة على الإنصاف والشمولية، ينبغي تدريب نظم الذكاء الاصطناعي التوليدي على البيانات التي تم تخليصها من التحيز والتي تمثل الأقليات المتأثرة.

وبالتالي يجب على المطورين والمستخدمين:

- اختبار نماذج الذكاء الاصطناعي التوليدي بعناية للتأكد من عدم تضمين أي تحيز في الأكواد أو الخوارزميات والتأكد من أن البيانات المستخدمة لتدريب الأداة خالية من التحيز وأنها تمثل مجموعات الأقليات.
- بذل الجهود لفهم البيانات المستخدمة لتدريب الأداة، ويجب أن يشمل هذا الفهم أصل البيانات ومحتوياتها وكيفية اختيارها وتجهيزها.
- تعزيز المعرفة بشأن التحيز والتنوع والشمول ومكافحة العنصرية ودعم القيم والأخلاق، لتحسين القدرة على اكتشاف المحتوى المتحيز أو التمييزي.

4.2 الموثوقية والسلامة

يضمن مبدأ الموثوقية والسلامة أن أنظمة الذكاء الاصطناعي التوليدي تلتزم بالمواصفات المحددة وفق الآلية التي كان يقصدها ويتوقعها المصممون، وتمثل الموثوقية مقياساً للثبات والثقة بمدى قوة الأنظمة، ومن ناحية أخرى تمثل السلامة مقياساً لضمان عمل أنظمة الذكاء الاصطناعي التوليدي على نحو آمن حيث لا يشكل خطراً على المجتمع والأفراد، وينبغي أن تكون له آليات مدمجة لمنع الضرر.

وبالتالي، يجب على المطورين والمستخدمين:

- تصميم وتطوير أنظمة الذكاء الاصطناعي التوليدي التي يمكنها تحمل حالات عدم اليقين وعدم الاستقرار والتقلبات التي قد تواجهها هذه الأنظمة.
- إنشاء مجموعة من المعايير والبروتوكولات لتقييم موثوقية النظام لتأمين سلامة الخوارزميات والمخرجات.
- وضع المشغلات/ التنبيهات المحددة مسبقاً بناءً على مخاطر النظام، ويجب أن تمتلك أنظمة الذكاء الاصطناعي التوليدي خيار تفعيل الرقابة البشرية.
- تعيين مختصين من أصحاب المصلحة للتحكم بالمشغلات/ التنبيهات كجزء من إجراءات تخفيف المخاطر أو التعافي من الكوارث التي تحتاج إلى الرقابة البشرية.
- تحديد المحتوى الذي تم إنشاؤه بالذكاء الاصطناعي التوليدي للتأكد من تنبيه المستخدمين بتحديات الموثوقية المحتملة، ويمكنهم التحقق من المحتوى عبر المصادر الأخرى.
- تشجيع المستخدمين على تحسين المحتوى الذي تم إنشاؤه بالذكاء الاصطناعي بمعلومات من مصادر موثوقة لضمان الدقة.
- السعي إلى فهم جودة وأصول بيانات التدريب التي يستخدمها نظام الذكاء الاصطناعي التوليدي لتعزيز موثوقية المحتوى.
- مراجعة المحتوى للتأكد من دقته والسياق العام لمنع نشر معلومات خاطئة.

4.3 الشفافية والقابلية للتفسير

من المهم بناء أنظمة وتقنيات الذكاء الاصطناعي التوليدي بدرجة عالية من الوضوح والقابلية للتفسير، مع وجود ميزات لتتبع مراحل اتخاذ القرارات المؤتمتة بطريقة مفهومة وواضحة. تعد الشفافية مهمة عند بناء الثقة في التفاعلات بين الإنسان والذكاء الاصطناعي التوليدي، إذ يساعد تميز المحتوى المعد من قبل أنظمة الذكاء الاصطناعي التوليدي على إزالة أي غموض مما يساعد المستخدمين على التمييز بين الاستجابات المؤتمتة والبشرية.

وبالتالي يجب على المطورين والمستخدمين:

- التواصل بوضوح عند استخدام الذكاء الاصطناعي التوليدي في التفاعلات مع الجمهور.
- إعلام المستخدمين عندما يتم إنشاء الرسائل أو المحتوى بالذكاء الاصطناعي.
- تقديم قنوات اتصال بديلة وغير مؤتمتة للمستخدمين الذين يفضلون التفاعلات البشرية.
- استخدام أدوات مثل العلامات المائية لتحديد المحتوى الناتج عن الذكاء الاصطناعي أو عن غيره.

4.4 المساءلة والمسؤولية

يتحمل المصممون والموردون والقائمون على المشتريات والمطورون ومقيمو أنظمة الذكاء الاصطناعي التوليدي المسؤولية الأخلاقية عن القرارات والإجراءات التي قد تؤدي إلى مخاطر محتملة وآثار سلبية على الأفراد والمجتمعات، وقد يترتب على اعتماد أنظمة الذكاء الاصطناعي التوليدي تبعات قانونية وأخلاقية مما يستدعي دراسة شاملة تشمل -على سبيل المثال لا الحصر- خطر انتهاك حقوق الملكية الفكرية والمخاوف بشأن حماية البيانات وحقوق الإنسان.

وبالتالي يجب على المطورين والمستخدمين:

- ضمان الحصول على البيانات وتصنيفها ومعالجتها وإتاحتها بشكل صحيح لتسهيل التدخل البشري والسيطرة عليها في مراحل لاحقة عند الحاجة.
- التأكد من فحوصات جودة البيانات وتنظيف البيانات والتحقق من سلامة البيانات من أجل الحصول على نتائج دقيقة.

- بناء النماذج والتحقق من صحتها بطريقة مسؤولة لتحقيق النتائج المرجوة.
- الالتزام بالتشريعات ذوات الصلة، مثل: نظام حماية البيانات الشخصية وأنظمة الملكية الفكرية لضمان الامتثال وحماية حقوق المستخدم.
- استشارة المتخصصين القانونيين لتقييم وتخفيف المخاطر المرتبطة باستخدام أنظمة الذكاء الاصطناعي التوليدي.

4.5 الخصوصية والأمن

من المهم تطوير واستخدام أنظمة الذكاء الاصطناعي التوليدي وتشغيلها بطرق تضمن حماية البيانات التي تجمعها، ومن ذلك المتطلبات النظامية المتعلقة بالأمن السيبراني لمنع انتهاك البيانات الذي يمكن أن يؤدي إلى إضرار بالسمعة أو ضرر نفسي أو مالي أو مهني أو أنواع أخرى من الضرر.

وبالتالي يجب على المطورين والمستخدمين:

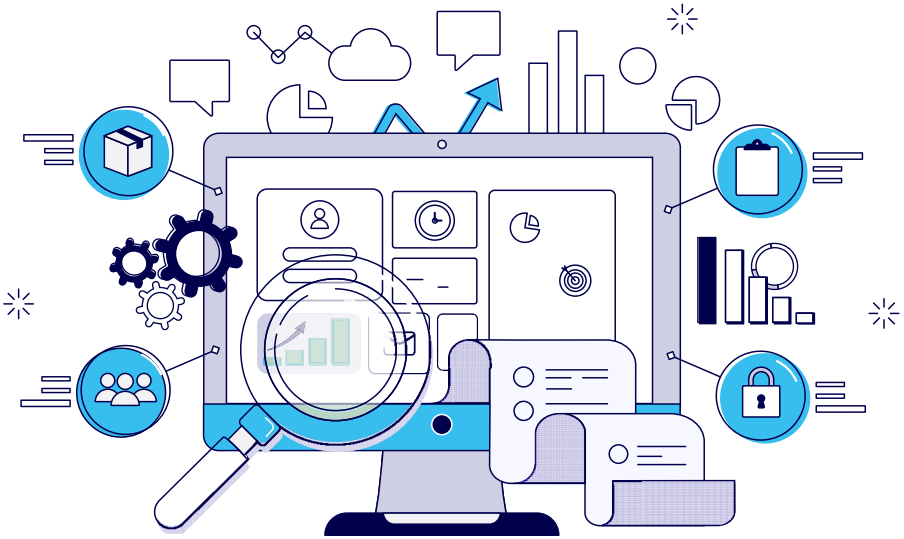
- التأكد من تطبيق تدابير الحماية والأمن الكافية عند استخدام البيانات المصنفة عند تدريب أنظمة الذكاء الاصطناعي التوليدي.
- تنفيذ تدابير صارمة لحماية البيانات والنظر في الأحكام المنصوص عليها في نظام حماية البيانات الشخصية وغيرها.
- تقييم المخاطر الناتجة عن استخدام نظام الذكاء الاصطناعي التوليدي وفقاً لمبادئ أخلاقيات الذكاء الاصطناعي: مخاطر بسيطة أو منعدمة، ومخاطر محدودة، ومخاطر عالية، ومخاطر غير مقبولة.
- يجب الاهتمام بتصميم النظام مع الأخذ بالاعتبار آليات الخصوصية والأمان وحماية الأبعاد المختلفة لنموذج الذكاء الاصطناعي وهيكليته ووحداته من التعديل غير المصرح به أو تلف أي من مكوناته.
- إعادة النظر بشكل مستمر في تقييم الأثر على الخصوصية وإدارة المخاطر لضمان تقييم الاعتبارات المجتمعية والأخلاقية بانتظام.

4.6 الإنسانية

ضرورة بناء واستخدام أنظمة الذكاء الاصطناعي باستخدام منهجية عادلة وأخلاقية تستند إلى حقوق الإنسان والقيم الثقافية الأساسية لإحداث أثر إيجابي على أصحاب المصلحة والمجتمعات في الأهداف والغايات الطويلة والقصيرة الأجل من أجل صالح البشرية، ومن الضروري أن يتم تصميم الأنظمة حيث لا تخدع أو تتلاعب أو تضع سلوكاً لا يقصد به تمكين المهارات البشرية أو تعزيزها أو زيادتها، بل ينبغي لها أن تتبنى نهجاً تصميمياً يركز على الإنسان ويتيح له الخيارات واتخاذ القرارات.

4.7 المنافع الاجتماعية والبيئية

يشمل هذا المبدأ الأثر المفيد والإيجابي للأولويات الاجتماعية والبيئية التي ينبغي أن تفيد الأفراد والمجتمع، مع التركيز على الأهداف والغايات المستدامة، ويجب ألا تتسبب أنظمة الذكاء الاصطناعي في الضرر أو تؤثر سلباً على البشر، بل يجب أن تساهم في تمكين واستكمال التقدم التقني والاجتماعي والبيئي مع السعي إلى معالجة التحديات المرتبطة بها.



5. مخاطر الذكاء الاصطناعي التوليدي وطرق تخفيفها

يخلق الذكاء الاصطناعي التوليدي فرصاً كبيرة، ولكن من المهم التأكد من استخدامه بطريقة آمنة وأخلاقية وقانونية، وينبغي إعطاء الأولوية لما يتعلق بالمخاطر والتدابير اللازمة لتخفيف حدتها.

5.1 التزييف العميق والتضليل

أدى الذكاء الاصطناعي التوليدي إلى تحول كبير في وسائل الإعلام الرقمي من خلال قدرته على إنشاء نصوص وصور وأصوات واقعية، ولكن يمكن أيضاً استخدام هذه الإمكانات لأغراض ضارة، مثل: عمليات الاحتيال المالي والابتزاز وسرقة الهوية، وفي كثير من الأحيان يستغل الأفراد البيانات من وسائل التواصل الاجتماعي والمنصات الأخرى لإنشاء شخصيات مزيفة (التزييف العميق)، مما يجعل التمييز بين المحتوى الحقيقي والمزيف أكثر صعوبة.

تدابير تخفيف المخاطر:

- استخدام العلامات المائية: يجب أن تتضمن جميع أنظمة الذكاء الاصطناعي التوليدي علامات مائية يمكن تحديدها، ويساعد هذا الإجراء الأفراد والجهات الأمنية على اكتشاف المحتوى المزيف العميق والمحتوى المُركب بالذكاء الاصطناعي والإبلاغ عنه دون التأثير على الاستخدام المشروع للذكاء الاصطناعي التوليدي.
- بروتوكولات "اعرف عميلك" (KYC) لموفري الخدمات السحابية: يجب على مقدمي خدمات القدرة الحاسوبية المكثفة والمستخدمة للذكاء الاصطناعي التوليدي ومقدمي وحدات معالجة الرسومات (GPUs) و/أو وحدات معالجة تنسر (TPUs)، واستخدام عمليات التحقق المشابهة لتلك التي تستخدمها المؤسسات المالية، على سبيل المثال: مطالبة المستخدمين بتقديم تفاصيل جواز السفر أو وثائق إثبات الهوية الأخرى لجعل الوصول إلى هذه الأدوات أكثر صعوبة على الجهات المحتالة.
- التحقق من المخرجات: يجب أن تُحلل أنظمة الذكاء الاصطناعي التوليدي المحتوى الذي تم إنشاؤه بحثاً عن أي محتوى غير لائق، مثل: خطابات الكراهية، وطلب تحويل الأموال، واستخدام وجوه الشخصيات العامة وأصواتهم.

- تعزيز المعرفة الرقمية والسلامة عبر الإنترنت: يجب أن توفر الشركات الخاصة والمؤسسات العامة تدريباً وحملات توعوية منتظمة تتضمن مخاطر التزييف العميق وأدوات وطرق التعرف عليها، بالإضافة إلى معلومات حول أهمية الحد من مشاركة البيانات الشخصية عبر الإنترنت، ومخاطر الاتصالات الرقمية التي لم يتم التحقق منها، وأهمية التحقق من طلبات البيانات من خلال قنوات بديلة موثوقة.

5.2 تهديدات السلامة

على الرغم من أن الذكاء الاصطناعي التوليدي يعد أداة فعالة للوصول إلى المعلومات في مجموعة واسعة من المواضيع، إلا أنه عند التلاعب به بنوايا غير حسنة، يمكن أن يؤثر سلباً على السلامة والأمن.

تدابير تخفيف المخاطر:

- أنظمة الإشراف وتصفية المحتوى: في الأنظمة الأساسية التي يمكن للمستخدمين الوصول إلى مخرجات الذكاء الاصطناعي التوليدي فيها، يعد نشر أنظمة قوية للإشراف على المحتوى وتصفيته أمراً بالغ الأهمية، ويجب أن تكون هذه الأنظمة قادرة على اكتشاف المحتوى الضار أو المسيء الناتج عن الذكاء الاصطناعي والإبلاغ عنه ومنع نشره، وإحدى الممارسات الجيدة هي التحقق من كل من طلب المستخدم ومخرجات النموذج، كما تضمن التحديثات والتحسينات المنتظمة لهذه الأنظمة حسب التغذية الراجعة والاتجاهات الناشئة أن تظل فعاليتها عالية ضد التهديدات المتطورة.
- تصفية مجموعة بيانات التدريب: يجب على مطوري الذكاء الاصطناعي التوليدي تقليل كمية المعلومات التي يحتمل أن تكون خطرة وتُستخدم في مجموعات البيانات التدريبية للنماذج، مع الأخذ في الاعتبار أن جميع المعلومات المستخدمة ستكون متاحة للعموم.
- الحد من الوصول المفتوح/ المصدر المفتوح لنماذج الذكاء الاصطناعي التوليدي العلمية: يجب على العلماء الذين يعملون على النماذج التي يمكن أن تخلق تهديدات للسلامة (مثل تخليق البروتين) ألا ينشروا مثل هذه النماذج للعامة دون مطالبة المستخدمين أولاً بالتحقق من هويتهم والأبحاث التي يعتزمون استخدام النموذج من أجلها.

5.3 المعلومات الخاطئة و"الهلوسة"

نظراً إلى طبيعة النماذج النصية للذكاء الاصطناعي التوليدي قد تكون بعض المعلومات التي تقدمها غير صحيحة، وفي بعض الأحيان يمكن أن تولد مثل هذه النماذج بثقة مفرطة "حقائق" تعد خيالاً كاملاً (تُعرف أيضاً باسم "هلوسة الذكاء الاصطناعي")، ودون مراجعة دقيقة لهذه المخرجات قد يتم نشر معلومات خاطئة.

تدابير تخفيف المخاطر:

- التحقق من المحتوى والاستشهاد: يجب أن تُقدر خدمات الذكاء الاصطناعي التوليدي المتاحة للعامة دقة المعلومات قبل عرضها على المستخدمين، ويجب أن تكون هذه الخدمات أيضاً قادرة على الاستشهاد بالمصادر التي تشير إلى مصدر المعلومة، وإذا لم يتمكن النموذج من إنتاج استجابة بمستوى عالٍ من الدقة وتحديد المصدر، فيجب أن تظهر رسالة محددة مسبقاً تحذر المستخدمين من أن المعلومات التي تم إنشاؤها قد لا تكون دقيقة.
- تصنيف المحتوى: يجب أن تتضمن أنظمة الذكاء الاصطناعي التوليدي علامات مائية في المخرجات بناءً على نوع الوسائط المستخدمة، ويمكن أن تتخذ العلامات المائية في الصور والفيديو شكل الطباعة المسبقة التقليدية أو في التشفير أو فك التشفير، ويمكن وضع علامة مائية على المحتوى الصوتي من خلال مقتطفات صوتية مضمنة أو تعديل الصدى، وفي النصوص يمكن استخدام مجموعات الكلمات والأحرف الخاصة أو خطوط غير قياسية.
- التحقق من المحتوى: تقع على عاتق المستخدمين مسؤولية التحقق من المحتوى الذي تم إنشاؤه بالذكاء الاصطناعي التوليدي، وهذا يعني أن على المستخدمين التدقيق في المحتوى والتحقق من صحته مقارنة بالمصادر الجديرة بالثقة، كما يجب على المستخدمين أيضاً البحث دائماً عن أي إشارة إلى أن يكون المحتوى قد جاء من أداة ذكاء اصطناعي توليدي، ففي حالة أن المحتوى بالفعل منتج بهذه الطريقة (انظر تصنيف المحتوى أعلاه).
- رفع الوعي: ينبغي تذكير المستخدمين بانتظام بأن مخرجات الذكاء الاصطناعي التوليدي يمكن أن تكون في بعض الأحيان غير دقيقة أو قديمة أو متحيزة أو حتى خادعة، ويجب تثقيفهم حول عمليات التحقق من الحقائق والمحتوى.

5.4 اختراق البيانات الحساسة

عند استخدام الذكاء الاصطناعي التوليدي يجب أن يدرك المستخدمون أنهم قد يكشفون عن معلومات حساسة عن غير قصد، لأن خدمات الذكاء الاصطناعي التوليدي تستخدم المعلومات الواردة من خلال الأوامر كبيانات تدريب لمواصلة تطوير النموذج ويمكن بعد ذلك كشفها لأطراف ثالثة، وذلك لأن نماذج الذكاء الاصطناعي التوليدي لا تخزن المعلومات في شكل مستند تقليدي، مما يعني أنه من الصعب تحديد ما إذا كانت أي معلومات حساسة قد تسربت بالفعل كما أنه من الصعب جعل النموذج "ينسى" هذه المعلومات.

تدابير تخفيف المخاطر:

- بروتوكولات استخدام الذكاء الاصطناعي التوليدي: يجب على الجهات تنفيذ سياسات لاستخدام الذكاء الاصطناعي التوليدي الذي يمنع المستخدمين من إدخال معلومات مصنفة في أدوات خارجية، ويجب أن تتضمن هذه التفاصيل الممارسات المقبولة في إنشاء المحتوى، والإجراءات اللازمة لمنع تسرب البيانات الحساسة، والخطوات اللازمة لدعم الاستخدام الأخلاقي للتقنية، وتعد المراقبة والمراجعات المنتظمة أمراً ضرورياً لتأكيد الالتزام بالسياسات.
- توعية الموظفين وتدريبهم: يجب على الجهات التأكيد على موظفيها أن المتطلبات النظامية الحالية والقادمة تظل ملزمة عند استخدام أدوات الذكاء الاصطناعي التوليدي، ويتضمن ذلك الالتزام بسياسات الجهة، بالإضافة إلى التنظيمات الأوسع التي تشمل حوكمة البيانات والأمن السيبراني، وتصنيف البيانات الحكومية، وحماية البيانات الشخصية، وحقوق الملكية الفكرية، والاعتبارات القانونية أو السياسية الأخرى ذات الصلة، ويمكن أن توفر الدورات التدريبية المستهدفة إرشادات حول كيفية تقليل المخاطر القانونية المحتملة المرتبطة باستخدام الذكاء الاصطناعي التوليدي.
- التحكم في بيانات المستخدم: يجب أن يوفر مزودو خدمات الذكاء الاصطناعي التوليدي للمستخدمين خيارات لمنح أو رفض الموافقة على استخدام بياناتهم لأغراض تدريب نماذج الذكاء الاصطناعي، ويجب عليهم أيضاً توفير خيار لإزالة سجل الأوامر بالكامل عند الطلب وفقاً للأنظمة واللوائح ذات الصلة.

5.5 الاحتيال في الشهادات

تعد عمليات إصدار الشهادات، مثل: الاختبارات والتقييمات المهنية بمثابة معايير أساسية للكفاءة الفردية والمصداقية المؤسسية، ومع ذلك مع ظهور الذكاء الاصطناعي التوليدي تواجه هذه العمليات تهديدات جديدة، ومن الممكن أن يساء استخدام الذكاء الاصطناعي التوليدي في قدرته على صياغة محتوى متخصص "شبيه بإنتاج الإنسان" لإنتاج إجابات أو مقالات أو حتى بحث تفصيلي، مما يزعزع المعايير التعليمية والمهنية التقليدية، ويشكل تقدم هذا المحتوى الناتج عن الذكاء الاصطناعي تحدياً واضحاً: فالمنتج المولد من قبل هذه الأنظمة أصلي ولم يتم نسخه مباشرة من مصادر معروفة مما يجعل من الصعب اكتشافه باستخدام أدوات مكافحة السرقة الأدبية التقليدية، وبالتالي فإن إساءة استخدام الذكاء الاصطناعي التوليدي دون رادع يمكن أن يعرض نزاهة التقييمات التعليمية والمهنية للخطر.

تدابير تخفيف المخاطر:

- تعزيز التقييم: مراجعة جميع أشكال التقييم لتعزيز صمودها ضد الاحتيال بمساعدة الذكاء الاصطناعي مع الأخذ بعين الاعتبار إجراء تعديلات، مثل: الانتقال إلى التقييمات الشخصية، أو مراجعة تنسيقات الأسئلة، أو اعتماد منهجيات تقييم مبتكرة، وعلاوة على ذلك يجب على الجهات إجراء تقييم شامل لامتحانات الشهادات، ولاسيما تلك الضرورية لشغل وظائف مثل: الوصول إلى البنية التحتية الحساسة أو تراخيص الممارسة الطبية، حيث تحدد هذه الشهادات أهلية الفرد في الحالات التي يمكن فيها استغلال الذكاء الاصطناعي التوليدي لتمرير مثل هذه الشهادات، ويجب على المؤسسات مراجعة عمليات الاعتماد بشكل استباقي لتقليل احتمالية استخدام الذكاء الاصطناعي التوليدي، وهذا يضمن أن التقييمات تحافظ على نزاهتها وأمنها خاصة في المجالات المهمة.
- التعليم والتدريب: إجراء تدريب شامل لرفع مستوى الوعي وتثقيف كل من المعلمين والطلاب حول الاستخدام المسؤول للذكاء الاصطناعي التوليدي، ويتضمن ذلك التعرف على سوء الاستخدام والالتزام بالممارسات الأخلاقية وتحديد السلوكيات المشبوهة التي قد تشير إلى الاحتيال.
- المبادئ التوجيهية والسياسات: تطوير وتنفيذ إرشادات واضحة بالتعاون مع الطلاب والمعلمين للتحكم في استخدام الذكاء الاصطناعي التوليدي في البيئات التعليمية، ويجب أن تتوافق هذه المبادئ التوجيهية مع أهداف التعلم وتحدد متى وكيف يمكن تطبيق الذكاء الاصطناعي التوليدي في التعليم.

- كفاءة الاستعلام: توفير تدريب متخصص لتعزيز كفاءة المعلمين والباحثين والطلاب في صياغة أوامر دقيقة وفعالة لأنظمة الذكاء الاصطناعي التوليدي للتأكيد على أهمية صياغة المدخلات التي تحقق مخرجات مثالية من الذكاء الاصطناعي.

5.6 انتهاك الملكية الفكرية والحماية

أدى ظهور الذكاء الاصطناعي التوليدي إلى تسليط الضوء على مسألة الاستخدام غير المصرح به أو تكرار المواد المحمية بحقوق الطبع والنشر، مما قد يؤدي إلى مسالات قانونية للملكية الفكرية للمحتوى الذي يتم إنشاؤه بالذكاء الاصطناعي، ويمتد هذا الخطر إلى الاستخدام غير المصرح به للمواد الحالية المحمية بحقوق الطبع والنشر وإنشاء محتوى جديد بالذكاء الاصطناعي قد ينتهك حقوق الملكية الفكرية عن غير قصد.

تدابير تخفيف المخاطر:

- ترخيص الملكية الفكرية والفحص النافي للجهالة: يجب على مطوري الذكاء الاصطناعي التوليدي الحصول على تراخيص لأي ملكية فكرية مدرجة في بيانات التدريب، مما يمنع كشط المحتوى العشوائي.
- يجب على العملاء إجراء الفحص النافي للجهالة الخاص بهم لتأكيد ما إذا كانت نماذج الذكاء الاصطناعي قد تم تدريبها على المحتوى المحمي قبل استخدامها.
- الحصول على إذن من المنشئ وآليات التعويض: يجب أن تحصل أنظمة الذكاء الاصطناعي التوليدي على إذن صاحب المحتوى الأصلي قبل إنشاء المحتوى بناءً على الملكية الفكرية الخاصة بها، ويجب أن تضع خدمات الذكاء الاصطناعي التوليدي آليات تعويض (مثل: أموال المساهمين) لمكافأة المبدعين بشكل عادل إذا تم استخدام ملكيتهم الفكرية في مجموعات التدريب على سبيل المثال، تقدم (Shutterstock) لمنشئي المحتوى خيار إلغاء استخدام أعمالهم في مجموعات التدريب على الذكاء الاصطناعي وأنشأت صندوقاً للمساهمين لتعويضهم عند استخدام أعمالهم.

5.7 تباين المخرجات

تعمل أنظمة الذكاء الاصطناعي التوليدي بشكل مختلف عن الخدمات التقليدية «المبرمجة» حيث يلتزم توليد المخرجات بخوارزميات محددة مسبقاً في جميع الحالات، وعلاوة على ذلك يمكن لمطوري هذه الخدمات تحديث عروضهم دون إبلاغ المستخدمين النهائيين، وبالتالي يجب أن يكون المستخدمون يقظين للغاية عند الاعتماد على المخرجات الناتجة عن الذكاء الاصطناعي التوليدي.

تدابير تخفيف المخاطر:

- توضيح للكود الذي تم إنشاؤه بالذكاء الاصطناعي التوليدي: يجب على مطوري الكود الذين يستفيدون من أدوات إنشاء كود عن طريق الذكاء الاصطناعي التوليدي، مثل (ChatGPT) و(GitHub Copilot) إضافة علامة توضيحية داخل الكود المصدري للإشارة إلى أنه تم إنشاؤه بالذكاء الاصطناعي، لا تعزز هذه العلامة التوضيحية الشفافية فحسب، بل يضمن أيضاً المسؤولية عن أصل الكود وجودته.
- التحقق بانتظام: يجب على المستخدمين التحقق بانتظام من صحة الكود الذي تم إنشاؤه بالذكاء الاصطناعي التوليدي والمحتويات الأخرى، خاصة في التطبيقات المهمة حيث تكون الدقة أمراً بالغ الأهمية، ويمكن أن يساعد تنفيذ عمليات الاختبار والتحقق الشاملة على تحديد وتصحيح أي تناقضات أو أخطاء في المخرجات.
- مواكبة التحديثات: للتخفيف من مخاطر التحديثات غير المعلنة لخدمات الذكاء الاصطناعي التوليدي، ويجب على المستخدمين مواكبة أي تغييرات أو تحسينات يتم إجراؤها من قبل مزودي الخدمة. يمكن أن يساعد الاشتراك في الإشعارات والتحديثات من المطورين على التكيف مع التعديلات في سلوك الخدمة أو جودة المخرجات.
- بناء الخبرة: يجب على المستخدمين وخاصة أولئك الذين يقومون بأدوار تطوير البرمجيات والترميز تطوير الخبرة في فهم الذكاء الاصطناعي التوليدي، وستساعد هذه الخبرة عند تقييم وتحسين المخرجات الناتجة عن الذكاء الاصطناعي على تلبية متطلبات محددة.

مايو 2025
نـــوع الوثيقـة | وثائق تنظيمية
تصنيف الوثيقـة | عـــام
رقـــم الوثيقـة | SDAIA-P115
رقـــم الإصـدار | 1

SDAIA.GOV.SA SDAIA_SA SDAIA.SAUDI SDAIA-KSA